Segmenting and Summarizing General Events in a Long-Term Lifelog

Yi Chen
CDVP, School of Computing
Dublin City University
Dublin, Ireland
ychen@computing.dcu.ie

Gareth J. F. Jones
CDVP & CNGL, School of Computing
Dublin City University
Dublin, Ireland
gjones@computing.dcu.ie

Debasis Ganguly
CNGL, School of Computing
Dublin City University
Dublin, Ireland
dganguly@computing.dcu.ie

ABSTRACT

Lifelogging aims to capture a person's life experiences using digital devices. When captured over an extended period of time a lifelog can potentially contain millions of files from various sources in a range of formats. For lifelogs containing such massive numbers of items, we believe it is important to group them into meaningful sets and summarize them, so that users can search and browse their lifelog data efficiently. Existing studies have explored the segmentation of continuously captured images over short periods of at most a few days into small groups of "events" (episodes). Yet, for long-term lifelogs, higher levels of abstraction are desirable due to the very large number of "events" which will occur over an extended period. We aim to segment a long-term lifelog at the level of general events which typically extend beyond a daily boundary, and to select summary information to represent these events. We describe our current work on higher level segmentation and summary information extraction for long term life logs and report a preliminary pilot study on a real long-term lifelog collection.

1. INTRODUCTION

Lifelogging uses digital devices to capture a person's life experiences. Current digital technologies are making it possible to record things one has seen or heard and to even detect what one was doing by analysis of one's digital activity records and sensor data. Examples of lifelog include [1,2]. Capture of an individual's lifelog can potentially last for many years. Over this time the lifelog might contain several years worth of video material, millions of personal images, many thousands of other files including emails, text messages, and various context data. It is unrealistic to expect people to easily browse such a vast collection of items. For this reason, applications such as that described in [3], are being developed to group certain types of data into small meaningful units, which are generally referred to as "events".

This type of segmentation can enable people to quickly scan for relevant sections of a lifelog covering a day, several days, a week or a longer period. However, according to [3], there are about 20 event per day, that is about 140 events a week, and more than 7000 events a year. Thus there would still be a very large amount of information to browse in a lifelog lasting several years. Thus, higher levels of segmentation and abstraction are desirable. Although grouping of events by dates, months and years can reduce the amount of items (events) that need to be displayed in a time period, it is unlikely that people always know the exact date associated with their required information. Further people do not necessarily want to browse their data using boxes defined by days, for example they may prefer to browse for a higher level "event", e.g. a holiday. We suggest that higher level segmentation should follow the way people remember their past experiences, so as to help them recognize which group (directory) they need to browse.

ECIR'10, Workshop on IAMPA, April 18, 2011, Dublin, Ireland.

In order to browse a lifelog collection, once items have been grouped into events, some form of surrogate summary is needed to represent the event. For example, a keyframe image may be selected from the event to help people recognize it. To enable people to recognize content associated with a certain activity based the information presented, it is important that the selected information is remindful enough to the user. Since the likelihood of recognizing the features of segments depends on how much they resemble the structures of the information in one's autobiographical memory, it is desirable that the segmentation algorithm can follow some general mechanisms of human autobiographical memory.

Autobiographical memory is the memory system responsible for the memory about one's individual's life. According to autobiographical memory theories [4], there are generally three levels of autobiographical memory: *lifetime periods*, *general events*, and *event-specific knowledge*. A lifetime event describes an extended period such as "when I was working at M company" or "when I was living in Y". A general event refers to a more specific period, which is usually in the form of a summary of repeated events of the same theme, such as "working on a small project", or an extended event like "a holiday in Italy". Event-specific knowledge usually contains vivid sensory-perceptual information of a specific event which happened in a consecutive time period (usually less than a day). Most of current event segmentation research has focused on this final level.

In the remainder of this paper, we report on a preliminary study to investigate the segmentation of lifelogs on the second level: general events, and an algorithm for the extraction of summary information for each general event.

2. PROTOTYPE DATA COLLECTION

It is essential to have a long-term lifelog data set to explore the effectiveness of general event level segmentation. As part of our ongoing lifelogging research, several participants have collected 20 months lifelog data collections containing items recording what they have seen and their activities at the computers, together with context information such as their location. The details of the data captured are suuarmarised as follows:

SenseCam photos: A wearable camera called Microsoft SenseCam was used to continuously capture what the lifelogger saw from the first personal perspective. The SenseCam proactively captures up to 6 images per minute.

Computer activities: Each window which comes to the foreground on a computer desktop was recorded by a software application called S'life. Information captured includes the title (name) of the window, the application which the window is opened by, the time when the window came to foreground and the time when it was closed. It also records the textual content and path of files in the window whenever applicable.

Context: The following context information was recorded:

Location: Location information was captured based on GPS and WiFi using Nokia 95 mobile phone. The location information is processed and stored in the form of five separated fields: country, country code, region (province, states, county), city, and street.

People: Names of Bluetooth devices near the lifelogger were captured by the Nokia N95 phone. It is expected that people name their Bluetooth device (e.g. mobile phones with Bluetooth) with their own names, so that the names of people near the lifelogger can be captured. In practice, not all people give their device a personal name or have their Bluetooth enabled at all times.

3. SEGMENTATION ALGORITHMS

Since general events usually take place over a period which contains repeated activities or share the same themes, we believe that it is possible to segment lifelog data to meaningful general events by detecting activity or theme distributions. We assume that in a general event, the information, attributes or features for the theme of the event occur densely, while such features may occur much less frequently in other parts of the lifelog (at least in its adjacent periods). Therefore, the segmentation of a lifelog into general events is similar to segmenting a textual document into several parts with differing main topics.

One method for segmenting textual documents into topical regions is the *TextTiling* algorithm originally developed by Hearst [5]. designed to segment expository texts, which are viewed as being composed of a sequence of main topics and a series of short, into subtopics regiosn. The TextTiling algorithm decomposes the text into blocks of a predefined size w; a pairwise comparison is made between the adjacent blocks using the standard vector space similarity measure, which greater similarity indicating a stronger match. Points with minimal similarity (i.e. valleys in the plotted graph of the similarities between the adjacent blocks) are chosen as the most likely candidates for segmentation boundaries. The valleys in the plotted graph are smoothed with a low pass filter (typically an averaging filter) before applying a threshold. In this preliminary long-term lifelog segmentation study, we explore segmentation based on computer activity information, location data and SenseCam images.

3.1 Segmentation of Computer Activities

We assume that people who spend a considerable time on their computers each day tend to have a periodic focus on computer activity themes. For example, in a certain 3 day period one may be interested in one topic, and read many items about it. Or one may spend some hours every day for several days working on a specific report. Of course, there are "gap" periods when the person is not focused on this topic or activity. We propose an approach of segmenting general events in (at least one aspect of) a person's lifelog by looking into the distribution of computer activities. The detailed procedure is as follows:

1. **Representing computer activities:** We use titles of active windows to represent computer activities. We assume that the window title is most repetitive for that activity.

2. **Creating a document:** A straightforward approach to creating a composite document is to merge all the window titles into a single string (document) in the order in which they occurred. However, such a document cannot fully represent the distribution of activities since it ignores the duration of activities. For example, one may spend 5 hours each day working on a document, but there may be only one record of this activity. At the same time, some other activities (e.g. loading their homepage on a web browser), which the person only spends 2 seconds on may be repeated many times each day. For this reason, we need to give long duration activities higher weight than short duration ones to reflect their importance. To this we repeat an item N times, where N=normalized duration of activity/normalized mean (if the duration of an activity is longer than the average).

3. **Segmenting documents:** We use Hearst's TextTiling algorithm [5] with a window size = 25 and smoothing parameter s = 1000.

3.2 Location based Segmenting algorithm

We assume that people may consider travelling or a holiday to be a general event in their lives. If they are the type of person who tends to have holidays in a place different from their regular location, these types of general events can be distinguished from others based on the location information.

3.2.1 Simplified approach

A straightforward approach to identifying such events is to read the location information one by one until it changes. As we described in section 2, location information in our database includes names of country, region, city and street names. Since street level location can change very frequently in short time periods (a few minutes or seconds), we excluded street level locations from this algorithm. City or region level change patterns can be variable. For example, some people may travel to an adjacent city or region to work every day. In this case, we consider that frequently occurring location names as "regular location", and only start segmenting when an unusual location name at the city level or above occurs. Details were as follows:

1) The most frequent city, region and country (which took more than 30% of the time) are extracted.

2) For each occurrence of a region or country which is different from these names, we start a new temporary segment. This segments ends when the region or country changes again.

3) Since when we are travelling it is possible that we pass by part of a region, city or country, if the duration of the segmentation is less than 2 hours, we did not consider that as a general event. Such segments are either joined with the previous segment or the latter. If the previous and latter segments share the same location information, the temporary segment is ignored.

3.2.2 TextTiling segmentation approach

Another approach to segmenting the lifelog based on location is to use the TextTilling method as we described above. We merge the location texts in the order of their timestamp with a format of: [country code][region][city][street]. Since the sample rate of capturing the location data is fixed, no additional repetition is needed. With this format of document, we anticipated that the segmentation could automatically detect stable locations for a given period. For example, a general event of having a holiday in France may involve frequent changes of street name, or even the city names, but the country level information would be stable (and is the most frequently occurring feature in this period), but different from the country names for the rest of time (the surrounding periods). Thus these stable location features could be used to distinguish this event from others.

3.3 SenseCam Concepts based Segmentation

We hypothesize that visual features may change in different general events. So we applied a similar document segmentation approach to segmenting a lifelog with the content of images, or more precisely, concepts identified content in the images, The application we used to detect the concepts in images is described in [6]. It returns a list of confidence scores for the presence of each of a set of 27 concepts typically found in SenseCam images in each image. We adopt the following procedure:

1) Sum the confidence scores for each concept for images in an event segmented by [3].
2) Merge the concepts (repeat N times) from all events in the lifelog to form a document, N = integer (the total confidence score of a concept in an event).
3) Segment with above TextTiling algorithm [5].

4. EVALUATION OF SEGMENTATIONS

In this pilot study, one of the three lifeloggers participated together with her lifelog collection. This includes about 450,000 images, 80,000 Slife records of about 2,000 hours of computer activities, and 18 months of context data (350,000 records). The latest data of the collection is about 14 months prior to the date of experiment. The data was segmented into four parallel sets:

S0) Segmented on weekly basis (baseline)

S1) Computer activity based segments

S2) Location based segments

S3) Simplified location based segments

S4) SenseCam concepts based segments (27 concepts)

The week based segmentation was used as a baseline partially because this time is the most straightforward segmentation base, but also because week, unlike month, is a perceptible temporal circle if a person distinguishes weekday and weekend. Yet, for some people, month may be a better segmentation base if they have more monthly events,

4.1 Method

A five point rating scale was used to evaluate each set of segmentation results (1=definitely a wrong segment point, to 5=definitely a correct segment point). Different materials were provided to assist the participant in rating the segmentations generated by the above algorithms. The average scores were compared with that of weekly baseline segmentation.

Judging Location-Based Segments: To rate the segmentation made based on location information, a list of merged location information was displayed with timestamps of the first and last captured records at that location before the location changes.

Judging Segmentation of Computer Activity Records: To assist the participant in judging the segmentation, we developed an experimental platform which shows:

1) daily computer activity records with: title and time of the top 5 activities in that day with the longest duration.
2) a list of segmentation points with timestamps.

Judging of Segmentation based on SenseCam Image Concepts: The DCU SenseCam browser[1] was used to assist the participant in recalling what was happening around the time of each segment point, so as to make judgments and ratings.

Judging Baseline Segments (Weekly): An experimental platform was developed based on the one used in judging computer activity based segments. Names of locations were added to the daily activity box.

4.2 Results

Five segmentation sets were evaluated using the data owner's manual judgment. Segments were considered to be bad if they were made between:

1) two identical locations: at city level or above

2) two identical computer activities: since we repeated the titles of long duration computer activities when creating the document for segmentation, it is possible that bad segmentation points can be placed between two of the same items which we repeated at document creation stage.

Table 1. Comparing segmentations results

	S0	**S1**	**S2**	**S3**	**S4**
Total segments	78	81	17	21	46
Bad segments	N/A	7	14	N/A	N/A
Average rating score	2.24	2.22	1.00	4.78	2.71

The simplified location-based approach (S3) was the most precise, while the TextTilling (S2) location-based segmentation did not get a single correct segment. In the latter case most of the segmentation points appeared between two identical records (with the same location at city level). The computer activity based and SenseCam concepts based segmentation approaches tend to have similar satisfactory scores as the weekly based segmentation, which is generally not very satisfactory. However, the score is entirely based on the participants subjective rating, which may not be very reliable, since the user may not always be able to tell what was happening based on the information provided when making the judgments.

5. SUMMARY GENERATION

After the events are segmented a remindful surrogate for the event needs to be generated or extracted to represent the segment, so as to help users decide which event to explore.

It has been found that salient items/events or things a person spent more effort on or repeated many times are better remembered. We hypothesize that activities or features, which occur frequently in one period, but less frequently in the remainder of the lifelog, are a good representative for that period. We developed summarizing algorithms based this hypothesis, and tested them through a self-rating scale regarding: how easily the participant could recognize the periods with the summary information.

5.1 Algorithms

The summary of each general event included the title of the main computer activity, a SenseCam image, and location information.

1. Summary Information from Computer activities

[1]http://sensecambrowser.codeplex.com/.

The representativeness score of a computer activity for a given period is calculated by: The total time of that activity during that period / total time of that activity during rest of the time in the lifelog collection. The top five highest score computer activities' titles of each general event were selected in the evaluation.

2. Key SenseCam image:

1) The key concept is calculated by the: (sum of the likelihood of the concept for images in the given period)/(sum of likelihood of the concept for images in the rest of the lifelog collection);

2) Key images were selected from the images in the period with highest likelihood for this concept.

3. Summary location information

The top two regions, countries, cities (if there were more than one of the above), and street names during the given period were selected to represent the location in that period.

5.2 Evaluation

5.2.1 Method

The summarizing algorithms were evaluated using the lifelog segmented into regions of general events. Since we did not get satisfactory results with the approaches in section 4, we decided to have the data manually segmented by the participant based on computer activities. This segmentation was combined with segmentations from location-based approach (S2). Summary information including key computer activity title, location and key SenseCam image was selected for each segmented general event.

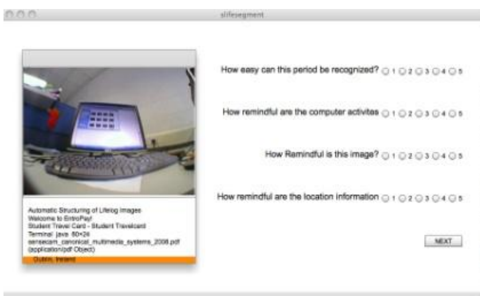

Figure1. Summary evaluation interface

An experimental platform was developed to present the summary information for the user to evaluate (Figure 1). A five point rating scale was used to investigate how easily a general event could be recognized based on the summary information, ranging from 1=unable to recognize the given period at all, to 5=it was extremely easy to recognize the general event with the summary information. Three five point rating scales were used for evaluating the effectiveness of each of the three types of summary information as memory cues for recognizing the periods, with 1=extremely ineffective to 5=extremely strong cue.

5.2.2 Results

The general easiness of recall effectiveness as memory cues for computer activities, location, and key SenseCam images achieved average rating scores of 3.6, 3.5, 1.9, and 2.8 respectively. The location information was particularly representative for the events when the person is away from routine location. Most of the selected computer activities were remindful for the activities that the user was doing, but they were not good cues for the exact time period, e.g. the year, month, etc. While some SenseCam images could be considered as representative for certain periods (mostly routine events), they are not good cues for recalling context information such as what time period an event was in. For general events which took place in a distinctive environment (e.g. a holiday abroad), although some key images were good cues, they did not usually concern the most representative scenes.

6. CONCLUSION

In this study, we examined automatic segmentation of general events in a prototype lifelog collection from an individual life logger. We segmented the lifelog using three types of data: records of computer activities, location, and visual concepts detected through content analysis of SenseCam images, and compared these with week based segmentation against a manual segmentation provided by the life logger. The main approach used for segmentation was to merge the text of records to form a single "document" representative, and then to segment it into sub-documents using the TextTilling algorithm. The results of the pilot study do not provide an immediate solution to the challenges of segmentation of long-term lifelogs. This may due to improper parameters of the TextTilling algorithm, the choice of the algorithm itself or the way in which documents were generated.

We also proposed an approach to selecting summary information to represent the events, and evaluated these using rating scales. We found that computer activity summaries were generally remindful regarding activities during the represented period, but they were not good cues for time attributes. Location provided good cues when it was distinctive from routine locations.

Further work will explore varying the segmentation parameters and alternative ways of creating the "document" to segment. Since how people segment temporal general events may be influenced by the context such as the current task and initial cues, future work may also explore dynamic segmentation which could cater for different tasks.

7. ACKNOWLEDGMENTS

This work is funded by a grant under the Science Foundation Ireland Research Frontiers Programme 2006, and supported by Microsoft Research under grant 2007-056.

8. REFERENCES

[1] Gemmell, J., Bell, G. and Lueder, R, 2006. MyLifeBits: A Personal Database for Everything. *Commun. ACM*.

[2] Lee, H., Smeaton, A., O'Connor, N., Jones, G., Blighe, M., Byrne, D., Doherty, A. and Gurrin, C. 2008. Constructing a SenseCam visual diary as a media process. *Multimedia Systems*, 14, 341-349.

[3] Doherty, A. R. and Smeaton, A. F. , 2008. Automatically Segmenting LifeLog Data into Events. In *Proceedings of 9th International Workshop on Image Analysis for Multimedia Interactive Services*

[4] Conway, M. A. and Pleydell-Pearce, C. W.,2000. The Construction of Autobiographical Memories in the Self-Memory System. *Psychol Rev*, 107, 2

[5] Hearst, M. A. ,2009. TextTiling: Segmenting Text into Multi-Paragraph Subtopic Passages. *Comput. Linguist.*, 23, 33-64.

[6] Byrne, D., Doherty, A. R., Snoek, C. G., Jones, G. J. F. and Smeaton, A. F. ,2009. Everyday Concept Detection in Visual Lifelogs: Validation, Relationships and Trends. *Multimedia Tools Appl.*, 49, 1, 119-144